Deploying Hadoop with SUSE® Manager

Big Data Made Easier

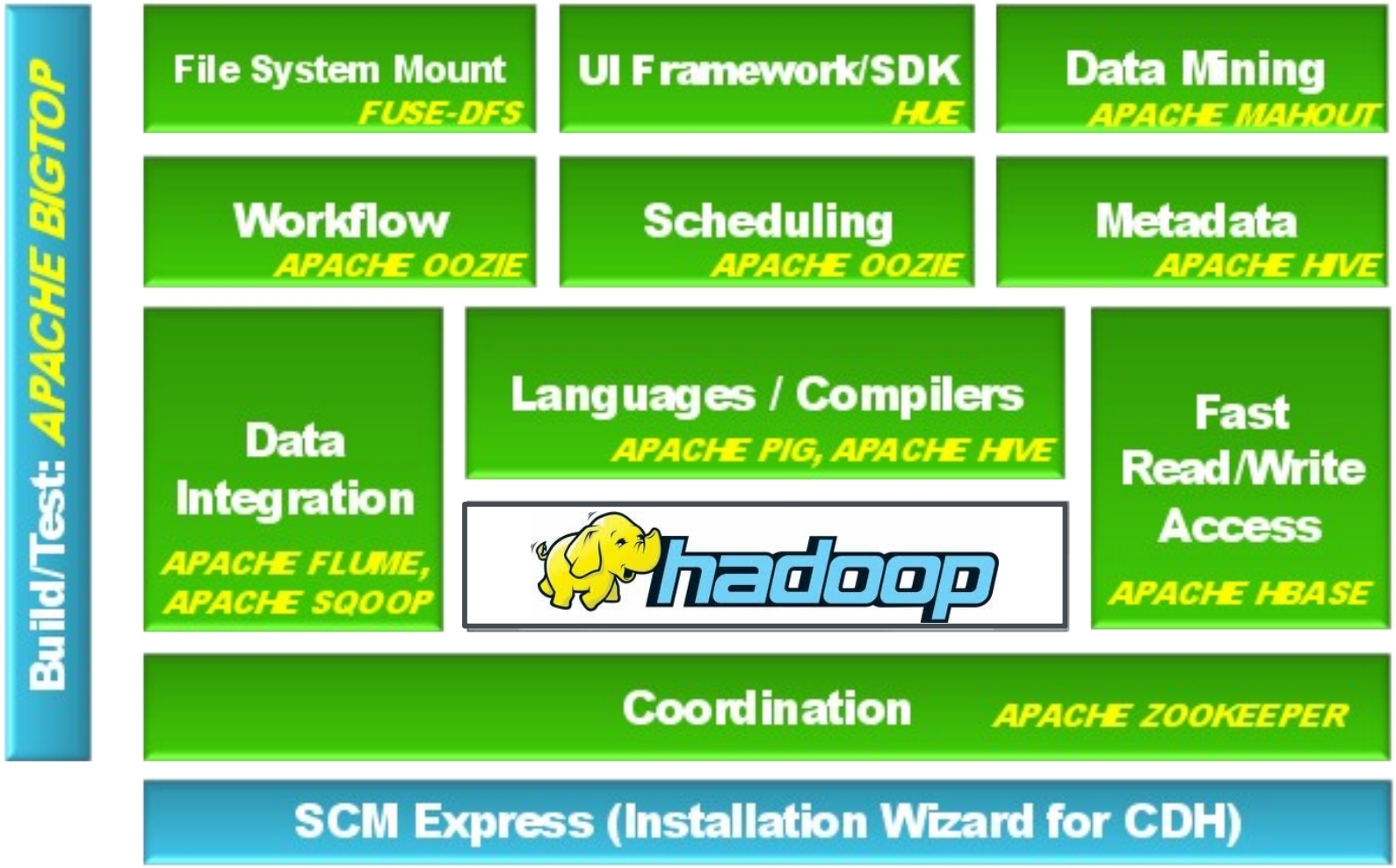
Peter Linnell / Sales Engineer
plinnell@suse.com

Alejandro Bonilla / Sales Engineer
abonilla@suse.com

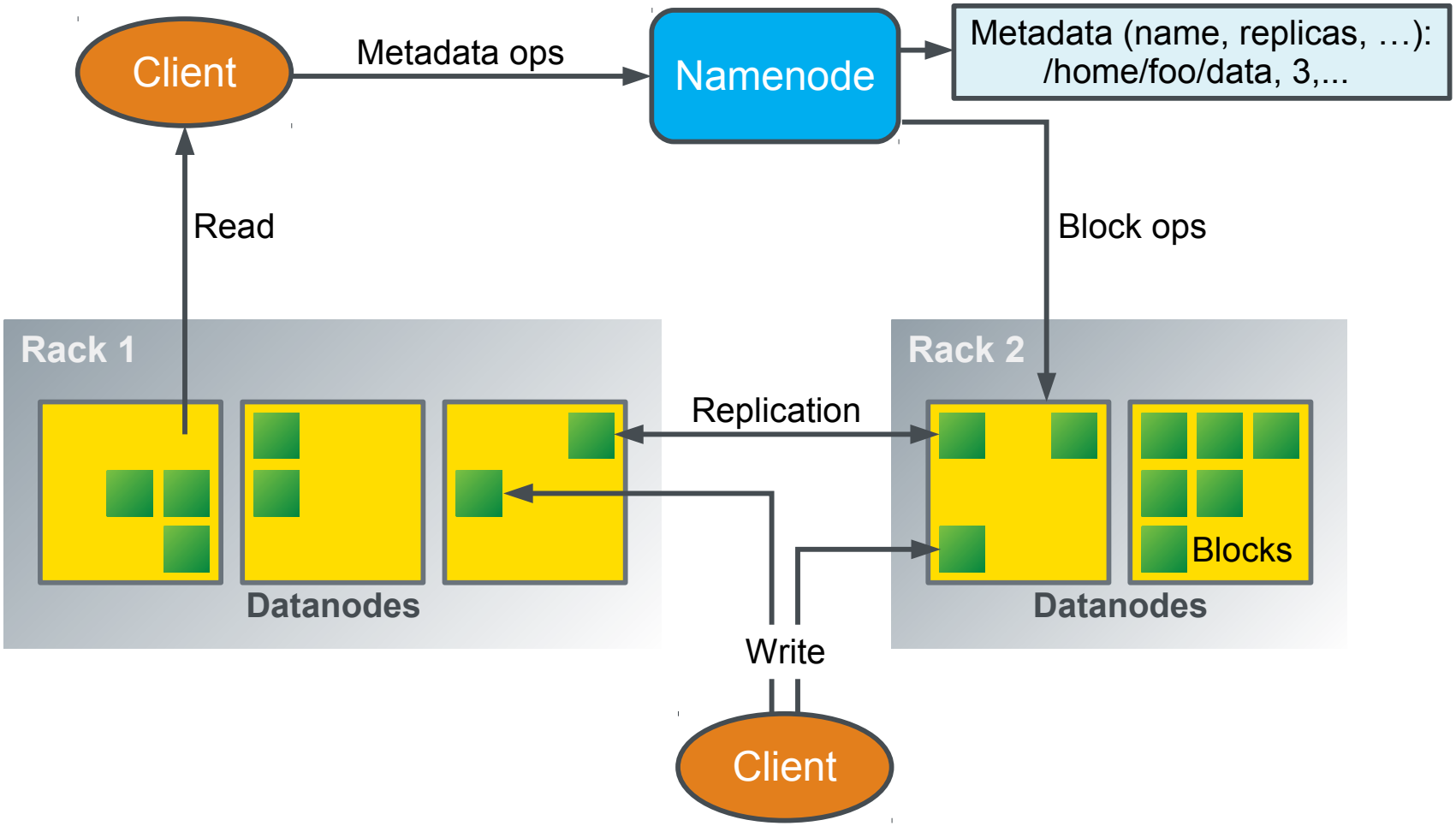
SUSE®
We adapt. You succeed.

Hadoop Core Components

Typical Hadoop Distribution



How Hadoop Works at Its Core



NameNode

- The NameNode (NN) stores all metadata
- Information about file locations in HDFS
- Information about file ownership and permissions
- Names of the individual blocks
- Location of the blocks
- Metadata is stored on disk and read when the NameNode daemon starts

NameNode2

- File name is fsimage
- Block locations are not stored in fsimage
- Changes to the metadata are made in RAM
- Changes are also written to a log file on disk called edits
- Each Hadoop cluster has a single NameNode
- The Secondary NameNode is not a fail-over NameNode
- The NameNode is a single point of failure (SPOF)

Secondary NameNode (master)

- The Secondary NameNode (2NN) is not-a fail-over NameNode!
- It performs memory/intensive administrative functions for the NameNode.
- Secondary NameNode periodically combines a prior file system snapshot and editlog into a new snapshot
- New snapshot is transmitted back to the NameNode
- Secondary NameNode should run on a separate machine in a large installation
- It requires as much RAM as the NameNode

DataNode

- DataNode (slave)
- JobTracker (master) / exactly one per cluster
- TaskTracker (slave) / one or more per cluster

Running Jobs

- A client submits a job to the JobTracker
- JobTracker assigns a job ID
- Client calculates the input and splits for the job
- Client adds job code and configuration to HDFS
- The JobTracker creates a Map task for each input split
- TaskTrackers send periodic “heartbeats” to JobTracker
- These heartbeats also signal readiness to run tasks
- JobTracker then assigns tasks to these TaskTrackers

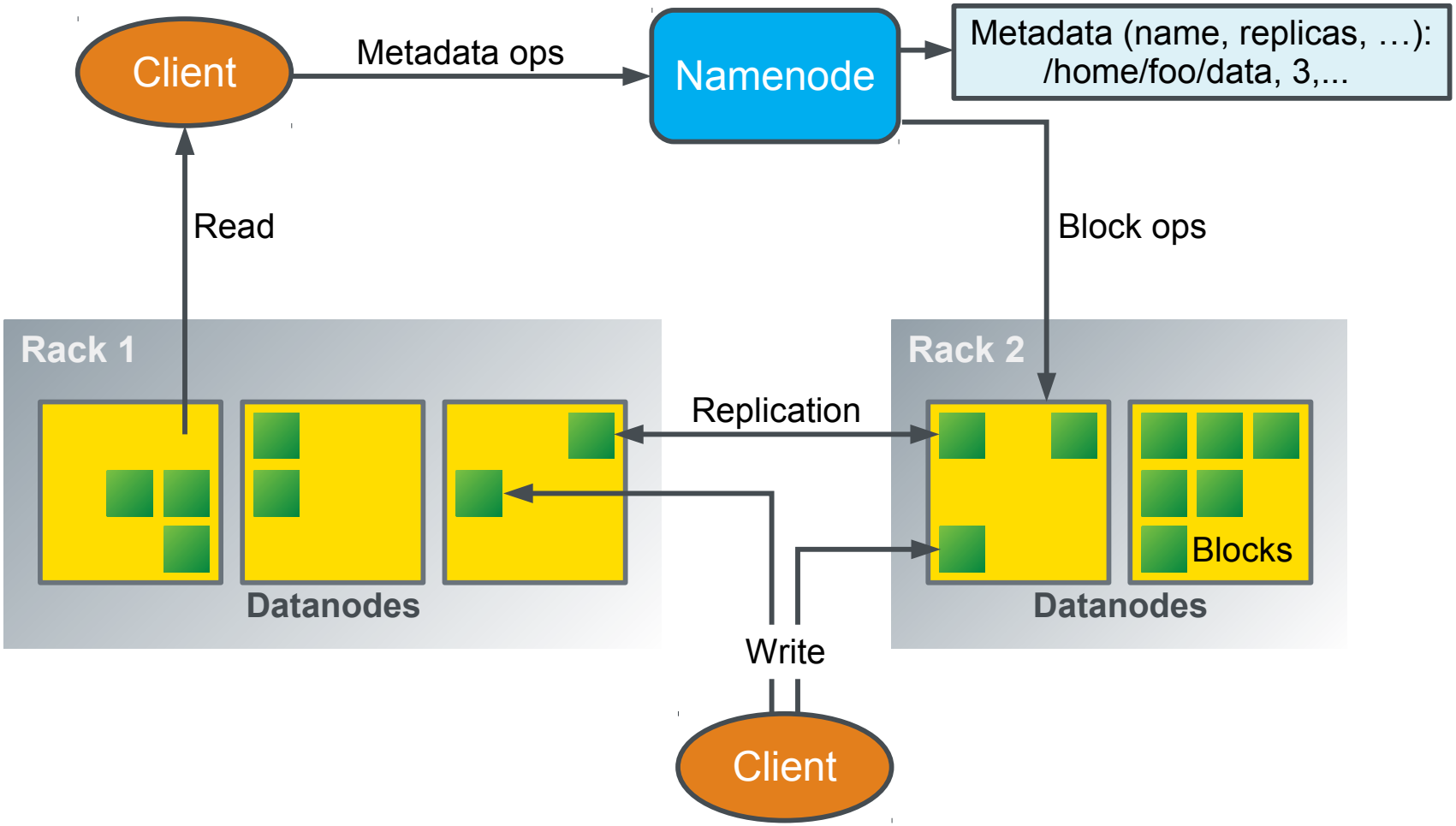
Running Jobs

- The TaskTracker then forks a new JVM to run the task
- This isolates the TaskTracker from bugs or faulty code
- A single instance of task execution is called a task attempt
- Status info periodically sent back to JobTracker
- Each block is stored on multiple different nodes for redundancy
- Default is three replicas

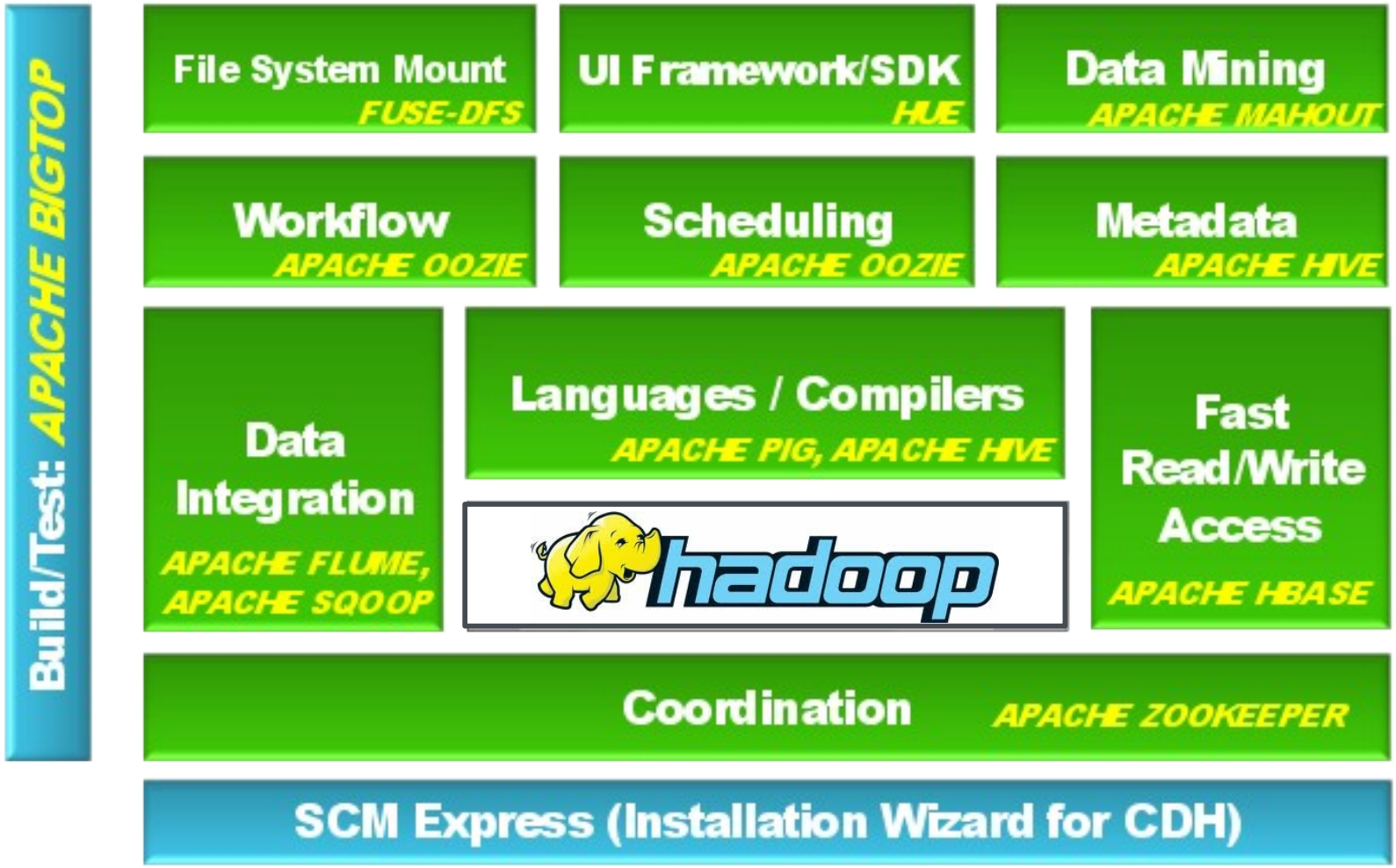
Anatomy of a File Write

1. Client connects to the NameNode
2. NameNode places an entry for the file in its metadata, returns the block name and list of DataNodes to the client
3. Client connects to the first DataNode and starts sending data
4. As data is received by the first DataNode, it connects to the second and starts sending data
5. Second DataNode similarly connects to the third
6. Ack packets from the pipeline are sent back to the client
7. Client reports to the NameNode when the block is written

Hadoop Core Operations – Review



Expanding on Core Hadoop



Hive, Hbase and Sqoop

Hive

- High level abstraction on top of MapReduce
- Allows users to query data using HiveQL, a language very similar to standard SQL

HBase

HBASE

- A distributed, sparse, column oriented data store

Sqoop

- The Hadoop ingestion engine – the basis of connectors like Teradata, Informatica, DB2 and many others.

Oozie

- Work flow scheduler system to manage Apache Hadoop jobs
- Workflow jobs are Directed Acyclical Graphs (DAGs) of actions
- Coordinator jobs are recurrent Workflow jobs triggered by time (frequency) and data availabilty
- Integrated with the rest of the Hadoop stack
  - Supports several types of Hadoop jobs out of the box (such as Java map-reduce, Streaming map-reduce, Pig, Hive, Sqoop and Distcp)
  - Also supports system specific jobs (such as Java programs and shell scripts)

Flume

- Distributed, reliable, and available service for efficiently collecting, aggregating, and moving large amounts of log data
- Simple and flexible architecture based on streaming data flows
- Robust and fault tolerant with tunable reliability mechanisms and many fail-over and recovery mechanisms
- Uses a simple extensible data model that allows for online analytic application

Mahout

- The Apache Mahout™ machine learning library's goal is to build scalable machine learning libraries
- Currently Mahout supports mainly three use cases:
  - Recommendation mining takes users' behavior and from that tries to find items users might like
  - Clustering, for example, takes text documents and groups them into groups of topically related documents
  - Classification learns from existing categorized documents what documents of a specific category look like and is able to assign unlabeled documents to the (hopefully) correct category

Whirr™

- Set of libraries for launching Hadoop instances on clouds
- A cloud-neutral way to run services
  - You don't have to worry about the idiosyncrasies of each provider.
- A common service API
  - The details of provisioning are particular to the service.
- Smart defaults for services
  - You can get a properly configured system running quickly, while still being able to override settings as needed

Giraph

- Iterative graph processing system built for high scalability
- Currently used at Facebook to analyze the social graph formed by users and their connections

Apache Pig

- Platform for analyzing large data sets that consist of a high-level language for expressing data analysis programs
- Language layer currently consists of a textual language called Pig Latin, which has the following key properties:
  - Complex tasks comprised of multiple interrelated data transformations are explicitly encoded as data flow sequences, making them easy to write, understand, and maintain.
  - Extensibility. Users can create their own functions to do special-purpose processing.

Ambari

- Project goal is to develop software that simplifies Hadoop cluster management
- Provisioning a Hadoop Cluster
- Managing a Hadoop Cluster
- Monitoring a Hadoop Cluster
  - Ambari leverages well known technology like Ganglia and Nagios under the covers.
- Provides an intuitive, easy-to-use Hadoop management web UI backed by its RESTful APIs

HUE – Hadoop User Experience

- Graphical front end to Hadoop tools for launching, editing and monitoring jobs
- Provides short cuts to various command line shells for working directly with components
- Can be integrated with authentication services like Kerberos or Active Directory
- More later on

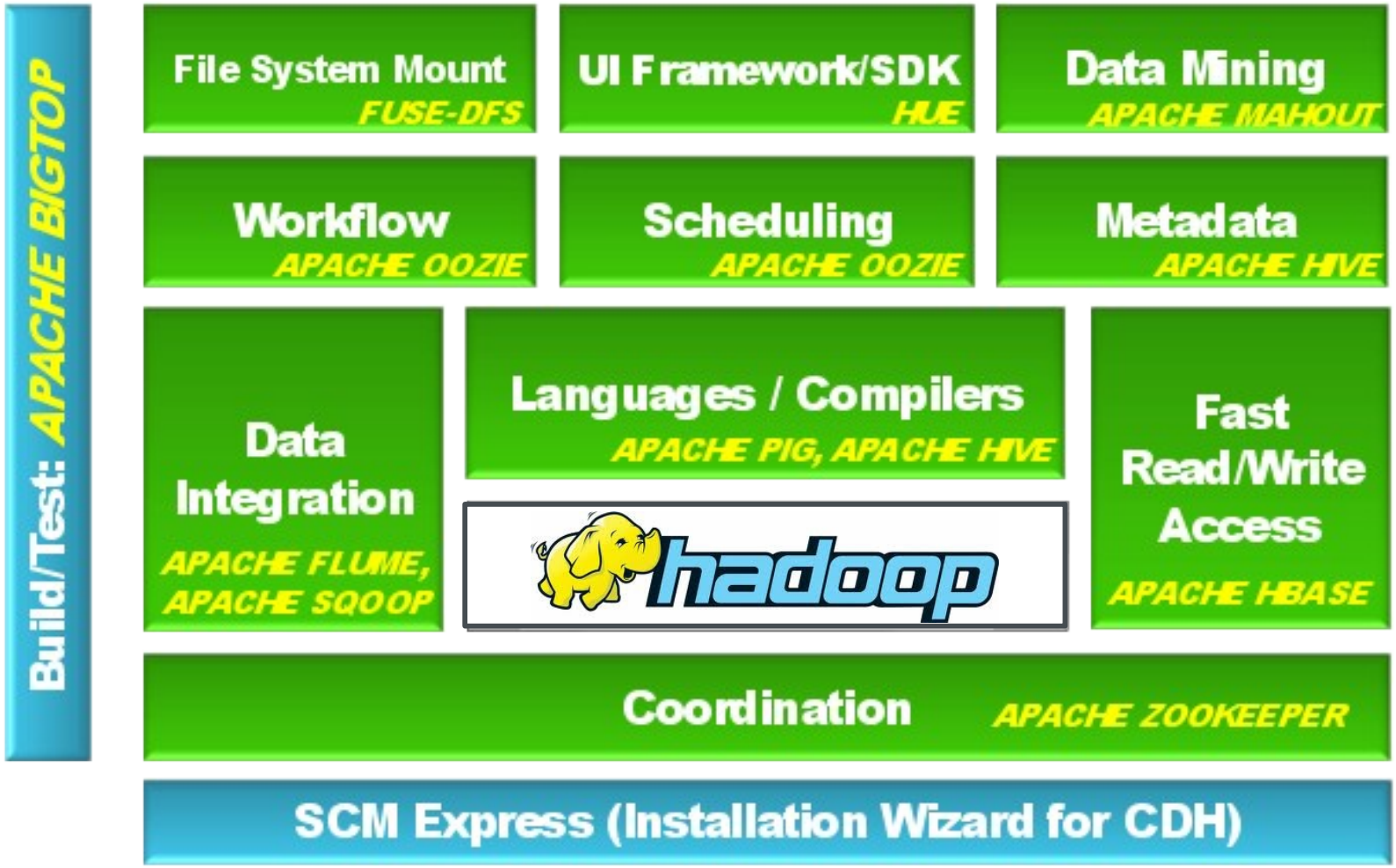
Zookeeper

- An orchestration stack.
- Centralized service for:
  - Maintaining configuration information
  - Naming
  - Providing distributed synchronization
  - Delivering group services.

Bigtop

- Packaging, QA testing and integration stack for Apache Hadoop components
- Made up of engineers from all the major Hadoop distros: Cloudera, Hortonworks, Intel and WANdisco, along with SUSE and independent contributors
- Almost unique among other Apache projects in that it integrates other projects as its goal
- All major Hadoop distros base their product on Bigtop

Typical Hadoop Distribution – Review



Web UI Ports for Users

- Daemon Default Port Configuration parameter
- NameNode 50070 dfs.http.address
- DataNode 50075 dfs.datanode.http.address
- Secondary NameNode 50090
  dfs.secondary.http.address
- Backup/Checkpoint Node 50105
  dfs.backup.http.address
- JobTracker 50030 mapred.job.tracker.http.address
- TaskTracker 50060 mapred.task.tracker.http.address

Why SUSE Manager ?

- Complements existing Hadoop Cluster Management Tools – None of which handle the OS stack
- Auto installation (Provisioning)
- Patch management
- Configuration management
- System and remote management (Groups)
- Monitoring

SHOWTIME!

Thank you.

Corporate Headquarters
Maxfeldstrasse 5
90409 Nuremberg
Germany

+49 911 740 53 0 (Worldwide)
www.suse.com

Join us on:
www.opensuse.org